Molecular Pharmaceutics 2018 vol.15 N10, pages 4378-4385

3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks

Kuzminykh D., Polykovskiy D., Kadurin A., Zhebrak A., Baskov I., Nikolenko S., Shayakhmetov R., Zhavoronkov A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 American Chemical Society. Convolutional neural networks (CNN) have been successfully used to handle three-dimensional data and are a natural match for data with spatial structure such as 3D molecular structures. However, a direct 3D representation of a molecule with atoms localized at voxels is too sparse, which leads to poor performance of the CNNs. In this work, we present a novel approach where atoms are extended to fill other nearby voxels with a transformation based on the wave transform. Experimenting on 4.5 million molecules from the Zinc database, we show that our proposed representation leads to better performance of CNN-based autoencoders than either the voxel-based representation or the previously used Gaussian blur of atoms and then successfully apply the new representation to classification tasks such as MACCS fingerprint prediction.

http://dx.doi.org/10.1021/acs.molpharmaceut.7b01134

Keywords

3D convolutional neural networks, autoencoders, wave transform, wavelets

References

[1] Aliper, A. M.; Plis, S. M.; Artemov, A. V.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. Mol. Pharmaceutics 2016, 13 (7), 2524-30, 10.1021/acs.molpharmaceut.6b00248

[2] Chen, L. Deep learning models for modeling cellular transcription systems, 2017.

[3] Kadurin, A.; Nikolenko, S. I.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. Mol. Pharmaceutics 2017, 14, 3098-3104, 10.1021/acs.molpharmaceut.7b00346

[4] Putin, E.; Mamoshina, P.; Aliper, A. M.; Korzinkin, M.; Moskalev, A.; Kolosov, A.; Ostrovskiy, A.; Cantor, C.; Vijg, J.; Zhavoronkov, A. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. Aging 2016, 8, 1021-33, 10.18632/aging.100968

[5] Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. Oncotarget 2017, 8, 10883, 10.18632/oncotarget.14073

[6] Makhzani, A.; Shlens, J.; Jaitly, N.; Goodfellow, I. Adversarial Autoencoders. International Conference on Learning Representations; San Juan, May 2-4, 2016.

[7] Todeschini, R.; Consonni, V.; Mannhold, R.; Kubinyi, H.; Folkers, G. Molecular Descriptors for Chemoinformatics: Vol. I: Alphabetical Listing/Vol. II: Appendices, References; Methods and Principles in Medicinal Chemistry; Wiley, 2009.

[8] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Proc. Edinburgh Math. SOC 1970, 1-14

[9] Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. J. Chem. Inf. Model. 1989, 29, 97-101, 10.1021/ci00062a008

[10] Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC international chemical identifier. J. Cheminf. 2015, 7, 23, 10.1186/s13321-015-0068-4

[11] Duvenaud, D. K.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. CoRR 2015, abs/1509.09292

[12] Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. J. Comput.-Aided Mol. Des. 2016, 30, 595-608, 10.1007/s10822-016-9938-8

[13] Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. CoRR 2013, abs/1312.6229

[14] Girshick, R. B. Fast R-CNN. 2015 IEEE Int. Conf. Computer Vision (ICCV) 2015, 1440-1448.

[15] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, 770-778.

[16] Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proc. 31st AAAI Conf. Art. Intelligence 2017, 4278

[17] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 2015.

[18] Krähenbühl, P.; Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. arXiv:1210.5644 [cs.CV] 2011.

[19] Pu, J.; Ramani, K. On visual similarity based 2D drawing retrieval. Computer-Aided Design 2006, 38, 249-259, 10.1016/j.cad.2005.10.009

[20] Zaharescu, A.; Boyer, E.; Varanasi, K.; Horaud, R. Surface feature detection and description with applications to mesh matching. 2009 IEEE Conf. Computer Vision and Pattern Recognition 2009, 373-380

[21] Johnson, A. E.; Hebert, M. Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes. IEEE Trans. Pattern Anal. Mach. Intell. 1999, 21, 433-449, 10.1109/34.765655

[22] Xiang, Y.; Mottaghi, R.; Savarese, S. Beyond PASCAL: A benchmark for 3D object detection in the wild. IEEE Winter Conference on Applications of Computer Vision Steamboat Springs, Mar 24-26, 2014; IEEE, 2014.

[23] Kazhdan, M. M.; Funkhouser, T. A.; Rusinkiewicz, S. Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. Symposium on Geometry Processing Aachen, Jun 23-25, 2003.

[24] Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D ShapeNets: A deep representation for volumetric shapes. 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 2015, 1912-1920

[25] Chang, A. X.; Funkhouser, T. A.; Guibas, L. J.; Hanrahan, P.; Huang, Q.-X.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; Xiao, J.; Yi, L.; Yu, F. ShapeNet: An Information-Rich 3D Model Repository. CoRR 2015, abs/1512.03012

[26] Maturana, D.; Scherer, S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2015, 922-928, 10.1109/IROS.2015.7353481

[27] Li, Y.; Pirk, S.; Su, H.; Qi, C. R.; Guibas, L. J. FPNN: Field Probing Neural Networks for 3D Data arXiv:1605.06240 [cs.CV] 2016.

[28] Riegler, G.; Ulusoy, A. O.; Geiger, A. OctNet: Learning Deep 3D Representations at High Resolutions. CoRR 2016, abs/1611.05009

[29] Engelcke, M.; Rao, D.; Wang, D. Z.; Tong, C. H.; Posner, I. Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks. 2017 IEEE International Conference on Robotics and Automation (ICRA) 2017, 1355-1361, 10.1109/ICRA.2017.7989161

[30] Shi, B.; Bai, S.; Zhou, Z.; Bai, X. DeepPano: Deep Panoramic Representation for 3-D Shape Recognition. IEEE Signal Processing Letters 2015, 22, 2339-2343, 10.1109/LSP.2015.2480802

[31] Chaudhuri, S.; Kalogerakis, E.; Guibas, L. J.; Koltun, V. Probabilistic reasoning for assembly-based 3D modeling. ACM Trans. Graph. 2011, 30, 1-10, 10.1145/2010324.1964930

[32] Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. CoRR 2013, abs/1312.6114

[33] Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A. A. Generative Adversarial Networks: An Overview. CoRR 2017, abs/1710.07035

[34] Wu, J.; Zhang, C.; Xue, T.; Freeman, B.; Tenenbaum, J. B. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling arXiv:1610.07584 [cs.CV] 2016.

[35] Nash, C.; Williams, C. K. I. The shape variational autoencoder: A deep generative model of part-segmented 3D objects. Comput. Graph. Forum 2017, 36, 1-12, 10.1111/cgf.13240

[36] Sharma, A.; Grau, O.; Fritz, M. VConv-DAE: Deep Vol.tric Shape Learning Without Object Labels; ECCV Workshops, 2016.

[37] Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. Applications of Deep Learning in Biomedicine. Mol. Pharmaceutics 2016, 13, 1445-1454, 10.1021/acs.molpharmaceut.5b00982

[38] Wallach, I.; Dzamba, M.; Heifets, A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. CoRR 2015, abs/1510.02855

[39] Gomes, J.; Ramsundar, B.; Feinberg, E. N.; Pande, V. S. Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity. CoRR 2017, abs/1703.10603

[40] Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. J. Chem. Inf. Model. 2017, 57 (4), 942-957, 10.1021/acs.jcim.6b00740

[41] Amidi, A.; Amidi, S.; Vlachakis, D.; Megalooikonomou, V.; Paragios, N.; Zacharaki, E. I. EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation. CoRR 2017, abs/1707.06017

[42] Torng, W.; Altman, R. B. 3D deep convolutional neural networks for amino acid environment similarity analysis. BMC Bioinf. 2017, 18, 302, 10.1186/s12859-017-1702-0

[43] Rafael Gonzalez, C.; Woods, E. R.; Eddins, L. S. Digital Image processing using MATLAB; Prentice Hall, 2003.

[44] Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions arXiv preprint arXiv:1610.02357 2016.

[45] Zeiler, M. D. ADADELTA: an adaptive learning rate method arXiv preprint arXiv:1212.5701 2012.

[46] Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. J. Chem. Inf. Comput. Sci. 2002, 42, 1273-1280, 10.1021/ci010132r